Recent Ensemble Smoother Applications: Data-Space Inversion and Deep Learning for Facies Models

Alexandre A. Emerick[1],
Mateus M. Lima[1], Carlos E.P. Ortiz[2],
Smith W.A. Canchumuni[3] and Marco Aurélio Pacheco[3]

[1]Petrobras, [2]UENF, [3]PUC-RIO

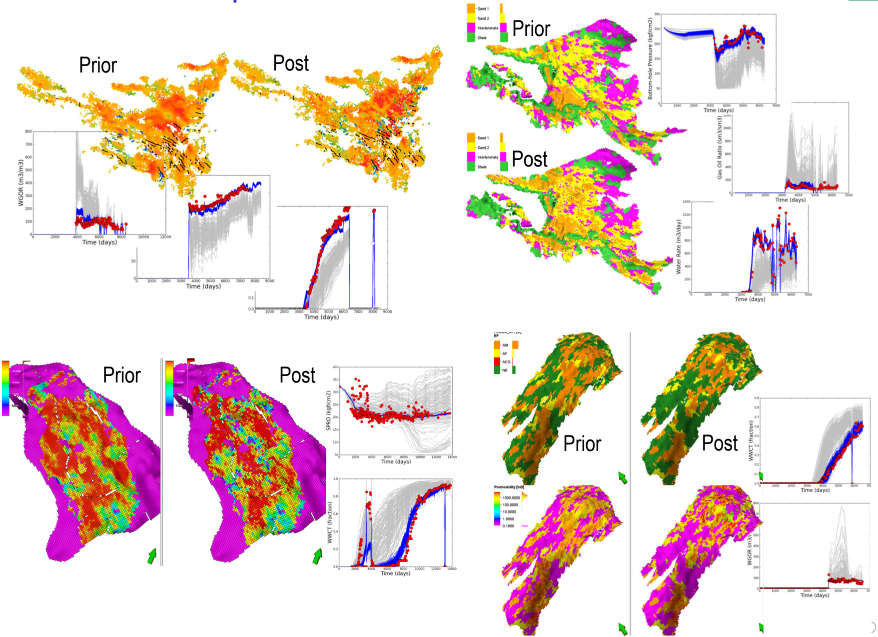
14th International EnKF Workshop
June 4, 2019

Outline

ES-MDA

- Reservoir history matching is a parameter-estimation problem.
- Simulation restarts required by EnKF (sequential data assimilation) are time-consuming:
  - Convergence problems in the reservoir simulations.
  - Very slow in clusters.
- ES is faster and simpler. But it does not match data sufficiently well.
- ES-MDA conciliates advantages of ES with better data matches: truly black-box and highly parallelizable.

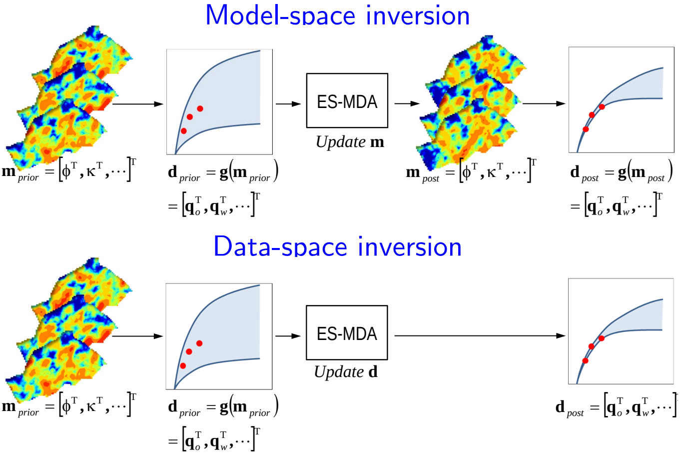
ES-MDA is in Operational Use



Outline

Data-Space Inversion*

- Inspired in the work from Sun and Durlofsky (2017)[1].
- Apply ES-MDA to updated directly the production profiles from a prior ensemble.

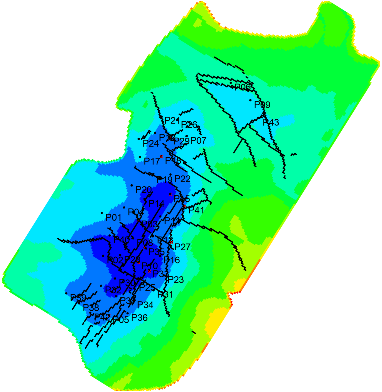


*Work with Mateus M. Lima (Petrobras) and Carlos E.P. Ortiz (UENF).

[1] Sun, W., Durlofsky, L., *A New Data-space Inversion Procedure for Efficient Uncertainty Quantification in Subsurface Flow Problems*, Math Geosci (2017).

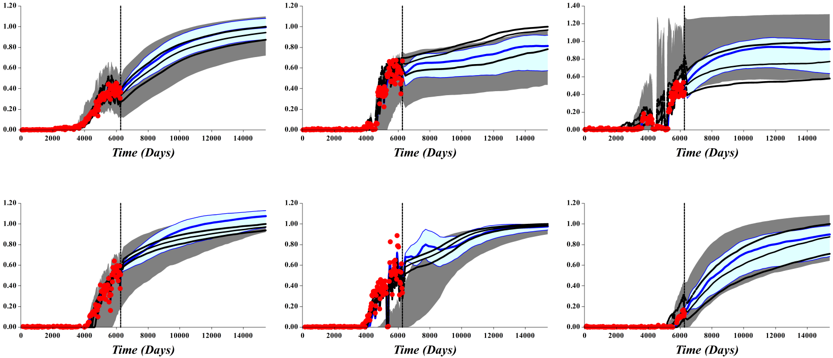
DSI – Field Case[2]

- Offshore turbidite Reservoir in Campos Basis.
- 18 years of production through 43 wells.
- Ensemble size: 500.
- Localization:
  - Space = 2 km.
  - Time = 6000 days.



[2] Lima, M.M.; Emerick, A.A. and Ortiz, C.E.P, *Data-Space Inversion with Ensemble Smoother*, arXiv:1903.09576 [math.NA] (2019).

Field Case – Water Production Rate



red dots: observed data

gray: prior (p10, p90)

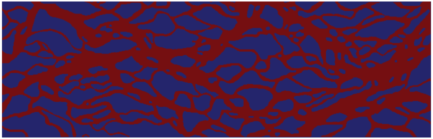
black: ES-MDA – model-space inversion (p10, p50, p90)

blue: DSI-ESMDA – data-space inversion (p10, p50, p90)

Outline

Facies Parameterization using Deep Learning*

- Updating channelized facies models is still a major challenge with ensemble data assimilation.



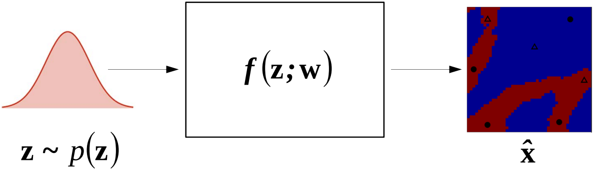
- Deep Learning (DL) emerged in the last decade as powerful technics for learning complex data representations[3].

*Work with Smith C. Arauco and Marco Aurélio Pacheco (PUC-Rio).

[3] Goodfellow, I., Bengio, Y. and Courville, A., *Deep Learning*, MIT Press (2016).

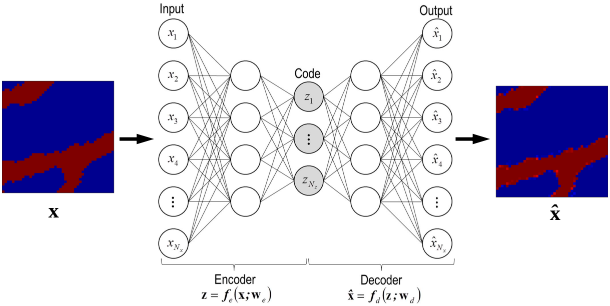
Generative Models

- Machine learning methods designed to generate samples from complex (and often with unknown closed form) probability distributions in high-dimensional spaces.
  - We would like to generate samples $\mathbf{x} \sim p(\mathbf{x})$.
  - We construct a deterministic function $\boldsymbol{f}(\mathbf{z};\mathbf{w})$ parameterized by $\mathbf{w}$, which receives a random argument $\mathbf{z} \sim p(\mathbf{z})$.
  - $\boldsymbol{f}(\mathbf{z};\mathbf{w})$ is modeled as a neural network, trained with a set of data points $\mathbf{x}_i$ such that if we provide $\mathbf{z} \sim p(\mathbf{z})$, it generates $\widehat{\mathbf{x}} \sim p(\mathbf{x}|\mathbf{z};\mathbf{w})$ which resembles samples from $p(\mathbf{x})$.



Autoencoders

- Combination of two neural networks:
  - Encoder: $\mathbf{z} = \boldsymbol{f}_e(\mathbf{x}; \mathbf{w}_e)$
  - Decoder: $\widehat{\mathbf{x}} = \boldsymbol{f}_d(\mathbf{z}; \mathbf{w}_d)$
- Training: find $\mathbf{w}_e$ and $\mathbf{w}_d$ that minimizes the reconstruction error, e.g., $\|\mathbf{x} - \widehat{\mathbf{x}}\|^2 = \|\mathbf{x} - \boldsymbol{f}_d(\boldsymbol{f}_e(\mathbf{x}; \mathbf{w}_e); \mathbf{w}_d)\|^2$.
- Autoencoders can be seen as nonlinear generalizations of PCA.



Variational Autoencoder (VAE)[4]

- For given training data, $\mathbf{x}$, we want to design a neural network by maximizing $p(\mathbf{x})$ with respect to a set of learnable parameters $\mathbf{w}$.

$$p(\mathbf{x}) = \int_{\mathbf{z}} \underbrace{p(\mathbf{x}|\mathbf{z})}_{\text{generative model}} \overbrace{p(\mathbf{z})}^{\text{prior}} d\mathbf{z} = \mathrm{E}_{\mathbf{z}\sim p(\mathbf{z})}\left[p(\mathbf{x}|\mathbf{z})\right].$$

- One alternative for training is to use Monte Carlo:
  - Sample a set of $\mathbf{z}_i \sim p(\mathbf{z})$ and compute $p(\mathbf{x}) \approx \frac{1}{N}\sum_i p(\mathbf{x}|\mathbf{z}_i)$.
  - Apply gradient ascent to maximize $p(\mathbf{x})$ with respect to $\mathbf{w}$.
- It won't work if $\mathbf{x}$ is high-dimensional (we need too many samples $\mathbf{z}_i$).
- Variational inference: introduce another (easy to sample) distribution $q(\mathbf{z}|\mathbf{x})$ and determine the parameters of q such that it generates samples $\mathbf{z}_i$ corresponding to high probability regions of $p(\mathbf{x}|\mathbf{z})$.

[4] Kingma, D.P. and Welling, M., *Auto-Encoding Variational Bayes*, arXiv:1312.6114 [stat.ML] (2013).

Variational Autoencoder (VAE)

- Instead of maximizing

$$p(\mathbf{x}) = \mathrm{E}_{\mathbf{z}\sim p(\mathbf{z})}\left[p(\mathbf{x}|\mathbf{z})\right]$$

we maximize

$$\widehat{p}(\mathbf{x}) = \mathrm{E}_{\mathbf{z}\sim q(\mathbf{z}|\mathbf{x})}\left[p(\mathbf{x}|\mathbf{z})\right]$$

- The trick is to use a well-known result from variational inference:

$$\overbrace{\ln p(\mathbf{x})}^{\text{log-evidence}} - \underbrace{\mathcal{D}_{\mathrm{KL}}\left[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z}|\mathbf{x})\right]}_{\text{error in the approximation}} = \underbrace{\mathrm{E}_{\mathbf{z}\sim q}\left[\ln p(\mathbf{x}|\mathbf{z})\right] - \mathcal{D}_{\mathrm{KL}}\left[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})\right]}_{\text{lower bound for the log-evidence}}$$

where $\mathcal{D}_{\mathrm{KL}}[q||p]$ is Kullback–Leibler divergence of q with respect to p.

- Instead of maximizing $\ln p(\mathbf{x})$ we can maximize its lower bound.

Variational Autoencoder (VAE)

- Mean field approach: select $q(\mathbf{z}|\mathbf{x}) = \mathcal{N}([\mu_1, \ldots, \mu_{N_z}]^\top, \mathrm{diag}[\sigma_1^2, \ldots, \sigma_{N_z}^2]^\top)$ and $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- During training, we minimize the loss function

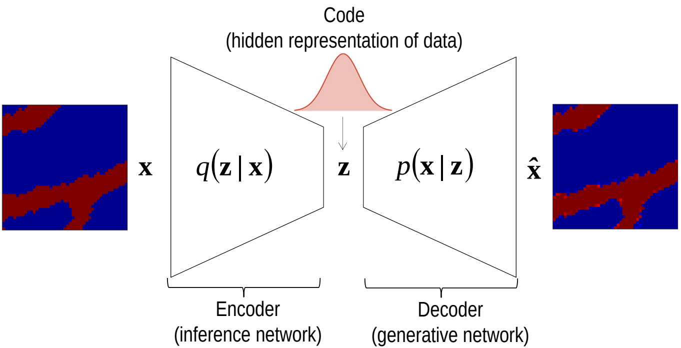
$$\mathcal{L}(\mathbf{x}) = \underbrace{\mathcal{L}_{\mathrm{RE}}(\mathbf{x})}_{\text{reconstruction error}} + \underbrace{\mathcal{D}_{\mathrm{KL}}\left[q(\mathbf{z}|\mathbf{x}) \| p(\mathbf{z})\right]}_{\text{regularization term}}$$

where

$$\mathcal{L}_{\mathrm{RE}}(\mathbf{x}) = -\frac{1}{N_x}\sum_{i=1}^{N_x}\left[x_i \ln(\widehat{x}_i) + (1 - x_i)\ln(1 - \widehat{x}_i)\right]$$

$$\mathcal{D}_{\mathrm{KL}}\left[q(\mathbf{z}|\mathbf{x}) \| p(\mathbf{z})\right] = \frac{1}{2}\sum_{i=1}^{N_z}\left(\mu_i^2 + \sigma_i^2 - \ln\left(\sigma_i^2\right) - 1\right)$$

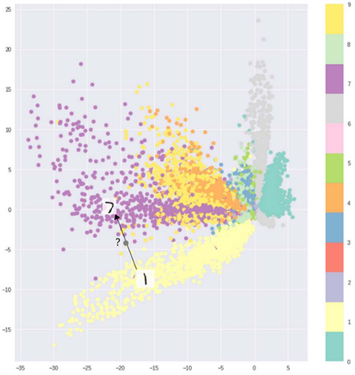
Variational Autoencoder (VAE)



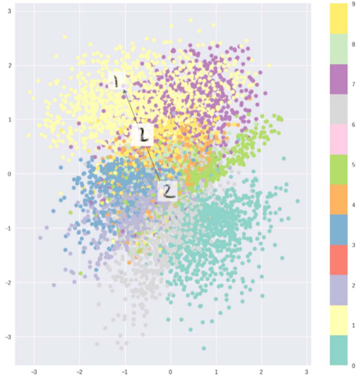
- $q(\mathbf{z}|\mathbf{x})$ encodes $\mathbf{x}$ in $\mathbf{z}$.
- $p(\mathbf{x}|\mathbf{z})$ decodes $\mathbf{z}$ in $\mathbf{x}$.
- Minimization of $\mathcal{L}_{\mathrm{RE}}(\mathbf{x})$ makes $\widehat{\mathbf{x}}$ to resemble $\mathbf{x}$.
- Minimization of $\mathcal{D}_{\mathrm{KL}}\left[q(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})\right]$ pushes $q(\mathbf{z}|\mathbf{x})$ to be similar to $\mathcal{N}(\mathbf{0},\mathbf{I})$.

- Example MNIST (handwritten digits) dataset[5,6].
- VAE generates a more continuous latent representation (easier to interpolate).

Autoencoder



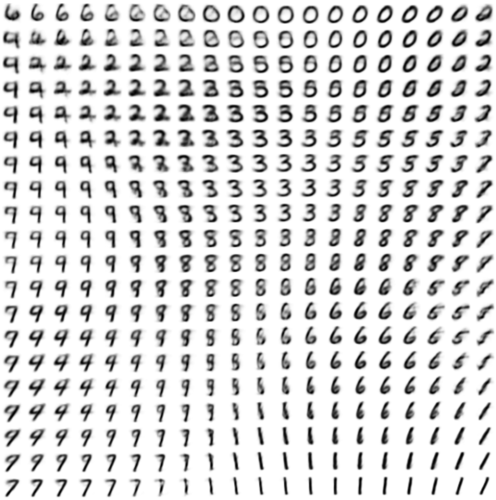
Variational autoencoder



[5] LeCun, Y., Cortes, C., and Burges, C.J.C., *The MNIST Database of Handwritten Digits*, `http://yann.lecun.com/exdb/mnist/`

[6] Shafkat, I., *Intuitively Understanding Variational Autoencoders*, `https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf`

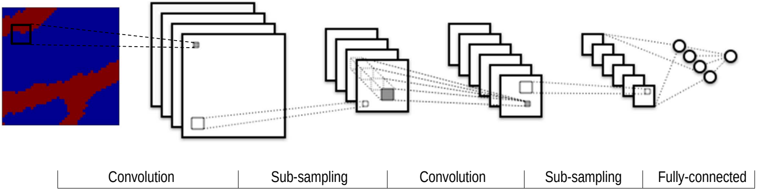
- Visualization of MNIST data with a two-dimensional latent space.
- VAE learned the data manifold[4].



[4] Kingma, D.P. and Welling, M., *Auto-Encoding Variational Bayes*, arXiv:1312.6114 [stat.ML] (2013).

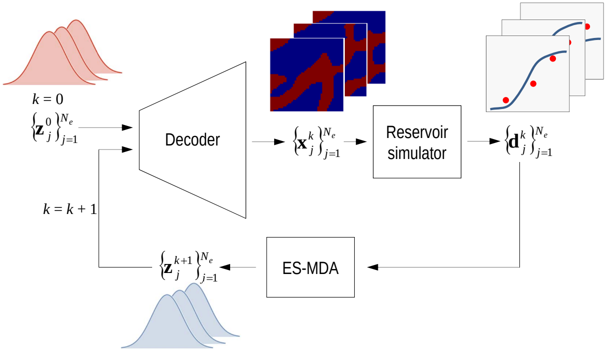
Convolutional Network[7]

- Standard fully-connected neural nets do not scale well for high-dimensional data $\mathbf{x}$.
- Convolution layers are specialized in data with grid structure such as images and time series.
- They provide a more efficient feature extraction by reducing the number of training parameters (weights in the filters).



[7] LeCun, Y. *Generalization and Network Design Strategies*. Tech report, University of Toronto (1989).

ES-MDA-CVAE[8]



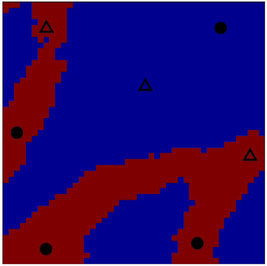
- After the CVAE is trained, we use ES-MDA to update the latent variables and the decoder to reconstruct facies.

[8] Canchumuni, S.W.A., Emerick, A.A. and Pacheco, M.A.C., *Towards a Robust Parameterization for Conditioning Facies Models Using Deep Variational Autoencoders and Ensemble Smoother*, Comput & Geosci (2019).

Test Case 1

- Two-facies model generated with *snesim*[9].
- 45×45 gridblocks.
- Constant permeability for each facies:
  - Channel: 5000 mD.
  - Background: 500 mD.

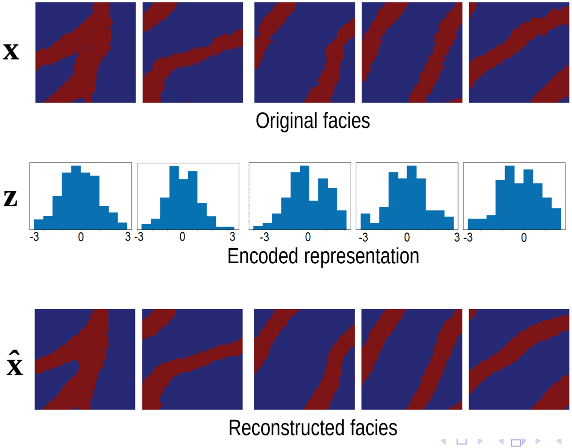
Reference



[9] Strebelle, S., *Conditional Simulation of Complex Geological Structures Using Multiple-point Statistics*, Math Geo (2002).

Test Case 1 – Training Results

- Training set: 24000 realizations. Validation set: 6000 realizations.
- 13 minutes in a cluster with four GPUs (NVIDIA TESLA P100).
- Reconstruction accuracy: 96.7%.

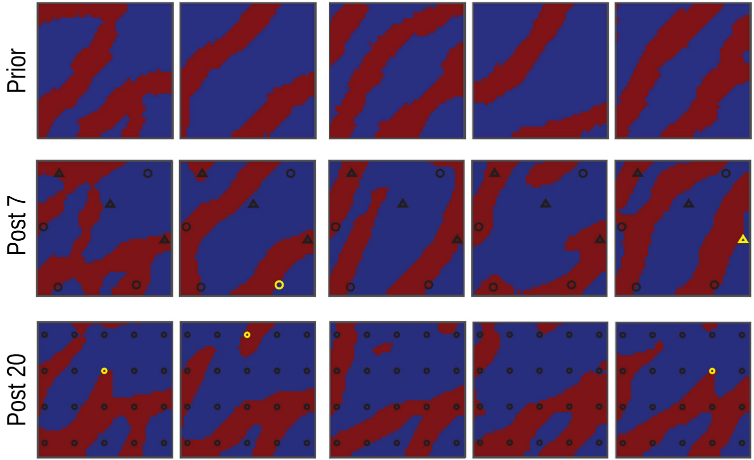


Test Case 1 – Testing the Decoder

- Generated a new realization by sampling $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
- Added small random perturbations: $\mathbf{z}^{k+1} = \mathbf{z}^{k} + \delta\mathbf{z}$, where $\delta\mathbf{z} \sim \mathcal{N}(\mathbf{0}, 0.1 \times \mathbf{I})$.

Test Case 1 – Data Assimilation (Facies Data)

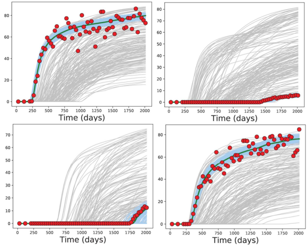
- MDA iterations: 4.
- Ensemble size: 200.

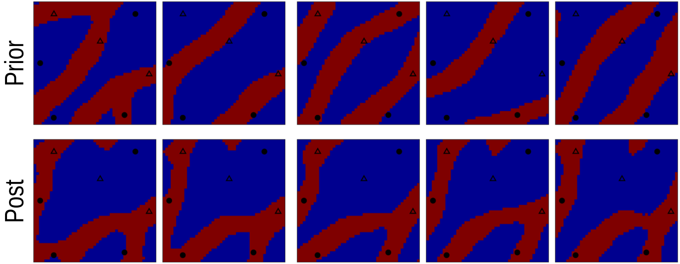


Test Case 1 – Data Assimilation (Production Data)

- Oil and water production data.
- MDA iterations: 4.
- Ensemble size: 200.

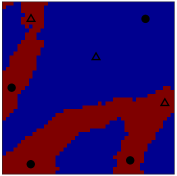
Water rate



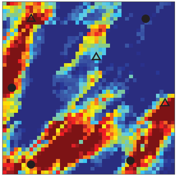


Test Case 1 – Comparison with Previous Results

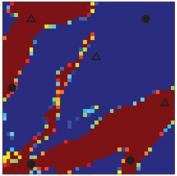
Reference



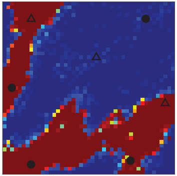
ES-MDA



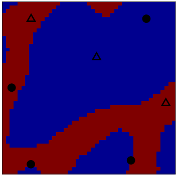
ES-MDA-OPCA[10]



ES-MDA-DBN[11]



ES-MDA-CVAE



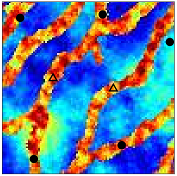
[10] Emerick, A.A., *Investigation on Principal Component Analysis Parameterizations for History Matching Channelized Facies Models with Ensemble-based Data Assimilation*, Math Geosci (2017).

[11] Canchumuni, S.W.A., Emerick, A.A. and Pacheco, M.A.C., *History Matching Channelized Facies Models using Ensemble Smoother with a Deep Learning Parameterization*, ECMOR (2018).

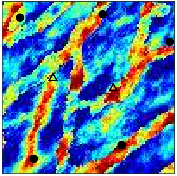
Test Case 2

- Two-facies model generated with *snesim*.
- 100×100 gridblocks.
- Simultaneous update of facies and permeability.
- Training: 32000, validation: 8000.
- Training time: 42 minutes*.
- Ensemble size: 200, MDA iterations: 20.

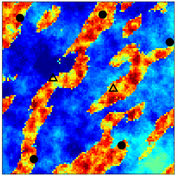
Reference



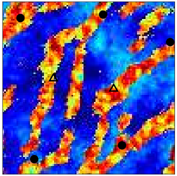
ES-MDA



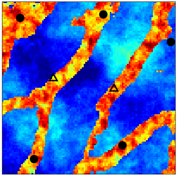
ES-MDA-OPCA



ES-MDA-DBN



ES-MDA-CVAE

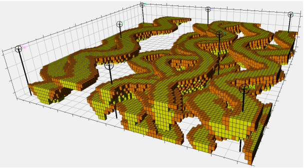


* Cluster with 4 GPUs (NVIDIA TESLA P100).

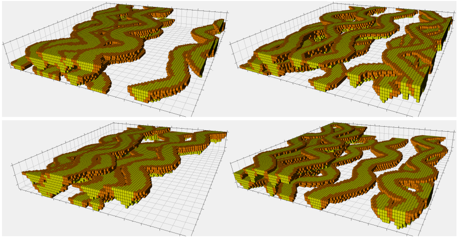
Test Case 3

- 3D channels (object-based simulation[12]).
- 3 facies: channel, levee and background.
- $100 \times 100 \times 10$ gridblocks.
- Training: 40000, validation: 10000.
- Training time: 49 hours*.
- Ensemble size: 200, MDA iterations: 20.

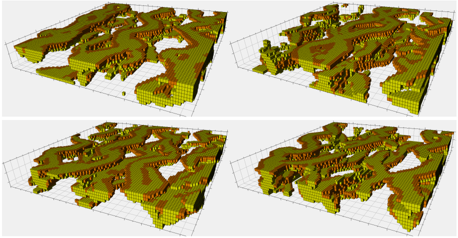
Reference



Prior



ES-MDA-CVAE

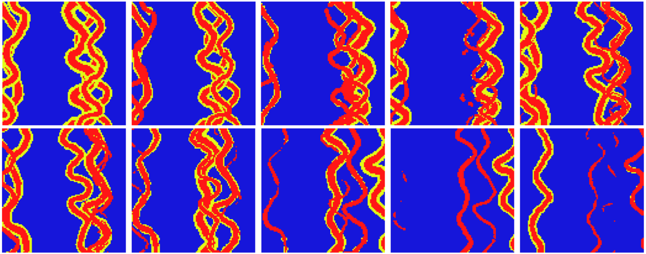


[12] Deutsch, C.V. and Journel, A.G. *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press (1998).

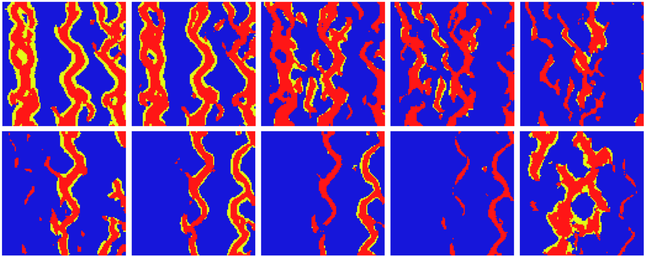
* Cluster with 4 GPUs (NVIDIA TESLA P100).

Test Case 3

Prior



ES-MDA-CVAE



Final Comments

- Data-space inversion:
  - Straightforward to apply.
  - It may serve as a first approximation.
  - It may be an useful for models with very complex geological description.
- Deep learning parameterization:
  - Promising results for facies models.
  - So far we tested only in small models ($\sim 10^4$–10^5 gridblocks).
  - It is unclear if it is going to be feasible in large-scale models ($> 10^6$ gridblocks).
  - Current implementation does not allow distance-based localization.

References I

Canchumuni, S. W. A., Emerick, A. A., and Pacheco, M. A.
History matching channelized facies models using ensemble smoother with a deep learning parameterization.
In *Proceedings of the 16th European Conference on the Mathematics of Oil Recovery (ECMOR XVI), Barcelona, Spain, 3–6 September*, 2018.
doi:10.3997/2214-4609.201802277.

Canchumuni, S. W. A., Emerick, A. A., and Pacheco, M. A. C.
Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother.
Computers & Geosciences, 128:87–102, 2019.
doi:10.1016/j.cageo.2019.04.006.

Deutsch, C. V. and Journel, A. G.
GSLIB: Geostatistical software library and user's guide.
Oxford University Press, New York, 2nd edition, 1998.

Emerick, A. A.
Investigation on principal component analysis parameterizations for history matching channelized facies models with ensemble-based data assimilation.
Mathematical Geosciences, 49(1):85–120, 2017.
doi:10.1007/s11004-016-9659-5.

Goodfellow, I., Bengio, Y., and Courville, A.
Deep learning.
MIT Press, 2016.
URL `http://www.deeplearningbook.org/`.

References II

Kingma, D. P. and Welling, M.
Auto-encoding variational Bayes.
arXiv:1312.6114 [stat.ML], 2013.
URL `https://arxiv.org/abs/1312.6114`.

LeCun, Y.
Generalization and network design strategies.
Technical report, University of Toronto, 1989.
URL `http://yann.lecun.com/exdb/publis/pdf/lecun-89.pdf`.

LeCun, Y., Cortes, C., and Burges, C. J. C.
The MNIST database of handwritten digits, 1998.
URL `http://yann.lecun.com/exdb/mnist/`.

Lima, M. M., Emerick, A. A., and Ortiz, C. E. P.
Data-space inversion with ensemble smoother.
arXiv:1903.09576 [math.NA], 2019.
URL `https://arxiv.org/abs/1903.09576`.

Strebelle, S.
Conditional simulation of complex geological structures using multiple-point statistics.
Mathematical Geology, 34(1):1–21, 2002.
doi:10.1023/A:1014009426274.

Sun, W. and Durlofsky, L. J.
A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems.
Mathematical Geosciences, Online, 2017.
doi:10.1007/s11004-016-9672-8.